Part 7
Lecture 1 Poisson Regression

1

Who I am…

Pascal Tyrrell, PhD

Associate Professor

Department of Medical Imaging, Faculty of Medicine

Institute of Medical Science, Faculty of Medicine

Department of Statistical Sciences, Faculty of Arts and Science

Poisson Regression

- The Poisson distribution describes the probability that a random event will occur in a time or space interval when the probability of the event occurring is very small, but the number of trials is very large.

- It is the limit of a binomial process in which:
  - $prob \rightarrow 0$
  - $n \rightarrow \infty$
  - n $* prob \rightarrow \mu$

Poisson regression models are generalized linear models with the Poisson distribution function.

The log link function is commonly used.

$$P(X = x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!}$$

Poisson Probability Distribution

- Our Poisson response variable may be modeled as:

$$Rate_j = \mu = \exp\{\beta_0 + \beta_1 X_j + \cdots + \beta_p X_p\}$$

Sometimes, the count responses will pertain to unequal units of time or space. In such cases, we let $\mu/t = \lambda$. [SAS: Use offset log(t)]

$$Rate_j = \lambda_j = \frac{\mu_j}{T_j} = \exp\{\beta_0 + \beta_1 X_j + \cdots + \beta_p X_p\}$$

❑ Using the Log Link, we obtain:

$$Rate_j = \lambda_j = \frac{\mu_j}{T_j} = \exp\{\beta_0 + \beta_1 X_j + \cdots + \beta_p X_p\}$$

$$log(Rate_j) = log(\lambda_j) = log\left(\frac{\mu_j}{T_j}\right) = \beta_0 + \beta_1 X_j + \cdots + \beta_p X_p$$

Overdispersion

- A characteristic of the Poisson distribution is that its mean is equal to its variance.
- If we see that the observed variance is greater than the mean - this is known as overdispersion. It tells us that the model is not appropriate.
- A common reason is the exclusion of relevant explanatory variables.

A SAS Example Using Poisson Regression:

Outcome: Number of patient visits to medical clinic (P_VISITS)

Predictors:
- ❑ Number of nearby housing units (HOUSING)
- ❑ Average household income (INCOME)
- ❑ Age of housing units (AGE_HOUSING)
- ❑ Distance of nearest clinic (NEAREST_CLINIC)
- ❑ Distance of clinic to nearby housing units (CLINIC_DISTANCE)

```
DATA poisson_expl ;
INPUT P_VISITS HOUSING INCOME AGE_HOUSING NEAREST_CLINIC CLINIC_DISTANCE @@ ;
DATALINES ;
 9   606   41393   3  3.04  6.32     10   392   36998   7  1.03  7.74     12   201   23864  43  4.80  8.74
 6   641   23635  18  1.95  8.89      0   828   85664   4  1.30  9.66     10   730   38647   9  0.67  7.92
28   505   55475  27  6.54  2.05     15   159   21238   4  2.98  8.66      8   738   58387  13  2.01  6.60
11   866   64646  31  1.67  5.81      9   830   47972  40  2.28  9.26      3   469   37242  40  1.42  8.37
 4   599   31972   7  0.72  8.11     16   234   33246  26  3.95  4.61     10   898   38337  32  2.63  9.56
 4   520   41755  23  2.24  6.81     29  1004   45927  24  4.90  2.69     10   780   68201   5  4.12  6.69
 0   354   46014  26  0.77  9.27      6   643   58315   8  0.78  6.26     15   622   41066  46  4.48  4.10
14   483   34626   1  3.51  7.92     26   741   69177   9  6.61  0.87      6   391   40873  19  1.67  6.90
16  1034   85207  13  4.23  4.40     13   306   40886  27  4.53  2.68      9   531   54655  40  2.32  5.69
13   456   33021  32  3.07  6.03      0   180   44588  14  0.88  9.38     21   566   49826   1  3.06  4.03
 9    19   39198  22  2.96  6.09      8   644   47347  35  2.94  7.69     13   410   29013  50  2.68  7.58
14   530   38794   5  2.77  6.08      8   109   31791   9  4.37  9.31      8   719   78082  31  2.70  4.89
 5   337   30855   1  1.33  9.86     21   809   42740  17  4.10  4.75      6   684   57506  51  2.13  8.31
 9   586   28852   7  2.98  8.64     12   722   59175  35  2.38  5.09      8   865   47118  46  2.17  9.06
 9  1113  120065   9  3.58  5.26     26  1006   48862  48  5.04  2.21     21  1031   72373  48  6.27  1.75
 7   525   32229   3  1.27  7.56      3   786   54678  20  3.59  8.52      7   862   67787   1  2.10  8.63
 4   377   36828  15  1.92  8.91      7  1041   59835  40  1.68  7.59     19   758   40305  15  3.95  5.58
26  1127   90302  26  5.83  1.74      5   524   51756  39  0.57  9.10     13  1141   50026  45  2.79  6.18
32   877   51707  27  5.19  3.66      9   725   34817  18  1.88  7.96     24  1289   98701   8  5.87  2.73
26  1007   89860  55  5.03  2.03     13   482   29942  14  3.17  6.91      7   674   58195  54  4.30  6.40
11   657   60513  32  4.38  8.30     28   666   68684  25  5.78  2.55      3   683   47991  57  1.54  9.52
12   302   42191  54  3.41  5.21     10   450   64790   3  4.35  6.03      8   650   63123  15  3.17  9.46
 3   603   28736  41  0.34  8.29     12   667   58535  25  2.78  5.59      9   406   39051  29  3.11  9.62
15   556   49129  33  4.78  3.89      6   921   42919  13  2.48  7.69     18   966  114633  38  6.33  2.22
12   635   29308  42  2.53  6.17     11   412   40722  32  2.47  9.43     12  1103   55773  44  4.58  8.68
 9   386   26734  14  4.99  9.70     12   526   42120  30  4.29  6.15     8    312   43393  41  2.25  6.43
14  1011   57862  54  4.60  3.94     11   523   28647  43  2.69  7.54     16   787   61765  53  5.39  3.37
10   925   70030  36  4.58  8.66      9  1066   61464  40  1.15  8.25      5   416   33348  48  1.48  7.66
22   898   46027  44  3.03  5.60      8  1001   70136  29  2.58  9.67      8   528   44541  31  4.91  9.67
 8   731   32202  43  5.15  9.67      9   669   34595  38  4.06  8.78     11   919   40795   8  2.97  7.79
 3   584   32871  13  1.47  8.02      8   582   30878  58  1.91  6.86     12   482   55972   9  2.91  5.85
11   439   29564  18  3.67  5.10      6   872   39366  52  0.73  8.67     14   781   33140  30  1.42  5.71
 2   153   46806  21  0.84  9.18      6   758   61563  31  3.08  8.33     17   120   19673  21  2.65  6.25
 6  1069   59805  22  2.50  9.43     15   782   38412  26  2.72  6.71     17   693   36190   6  4.70  9.54
11   443   42555  53  2.62  5.75     15   551   41045   2  3.62  7.45      6   348   25768  42  1.43  7.11
15   780   53974  47  4.21  6.41     10   752   71814   1  3.13  5.47      6   817   54429  47  1.90  9.90
 4   268   34022  54  1.20  9.51     6   519   52850  43  2.92  8.62
; RUN ;
```

```
PROC GENMOD DATA = POISSON_EXPL;
MODEL P_VISITS = HOUSING INCOME AGE_HOUSING NEAREST_CLINIC CLINIC_DISTANCE  / DIST =
POISSON LINK   = LOG;
OUTPUT OUT=TEMP P=MUHATI RESDEV=DEVI;
RUN;

PROC PRINT DATA = TEMP (OBS=10);
VAR P_VISITS MUHATI DEVI;
RUN;

DATA TEMP;
SET TEMP;
ID = _N_;
RUN;

SYMBOL1 V=DOT I=JOIN C=BLUE H = .8;
AXIS1 LABEL=(ANGLE = 90);

PROC GPLOT DATA = TEMP;
PLOT DEVI*ID/ VAXIS = AXIS1;
RUN;
QUIT;
```

An Interpretation of Parameter for Clinic Distance:

If the distance from the clinic was to increase by a kilometer, the difference in the logs of expected counts of patients visiting the clinic would decrease by 0.1288, while holding all other variables in the model constant.

An Interpretation for Clinic Distance:

If the distance from the clinic were to increase by one kilometer, the number of patients visiting the clinic would decrease by about 1 patient (0.879), while holding all other variables in the model constant.

My note: exp(- 0.1288) = 0.879

The GENMOD Procedure

Model Information	
Data Set	WORK.POISSON_EXPL
Distribution	Poisson
Link Function	Log
Dependent Variable	P_VISITS

Number of Observations Read	110
Number of Observations Used	110

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	104	114.9854	1.1056
Scaled Deviance	104	114.9854	1.1056
Pearson Chi-Square	104	101.8808	0.9796
Scaled Pearson X2	104	101.8808	0.9796
Log Likelihood		1898.0224	
Full Log Likelihood		-279.5121	
AIC (smaller is better)		571.0243	
AICC (smaller is better)		571.8398	
BIC (smaller is better)		587.2272	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	2.9424	0.2072	2.5362	3.3486	201.57	<.0001
HOUSING	1	0.0006	0.0001	0.0003	0.0009	18.17	<.0001
INCOME	1	-0.0000	0.0000	-0.0000	-0.0000	30.63	<.0001
AGE_HOUSING	1	-0.0037	0.0018	-0.0072	-0.0002	4.37	0.0365
NEAREST_CLINIC	1	0.1684	0.0258	0.1179	0.2189	42.70	<.0001
CLINIC_DISTANCE	1	$\boxed{-0.1288}$	0.0162	-0.1605	-0.0970	63.17	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

Calculation of Deviance

First 10 Fitted Values & Deviance Residual

The SAS System

Obs	P_VISITS	MUHATI	DEVI
1	9	12.3378	-0.99880
2	10	6.6737	1.19816
3	12	10.0527	0.59580
4	6	8.7671	-0.99158
5	0	4.0673	-2.85212
6	10	7.3332	0.93268
7	28	28.1259	-0.02375
8	15	8.6908	1.93784
9	8	8.5615	-0.19406
10	11	8.4071	0.85335

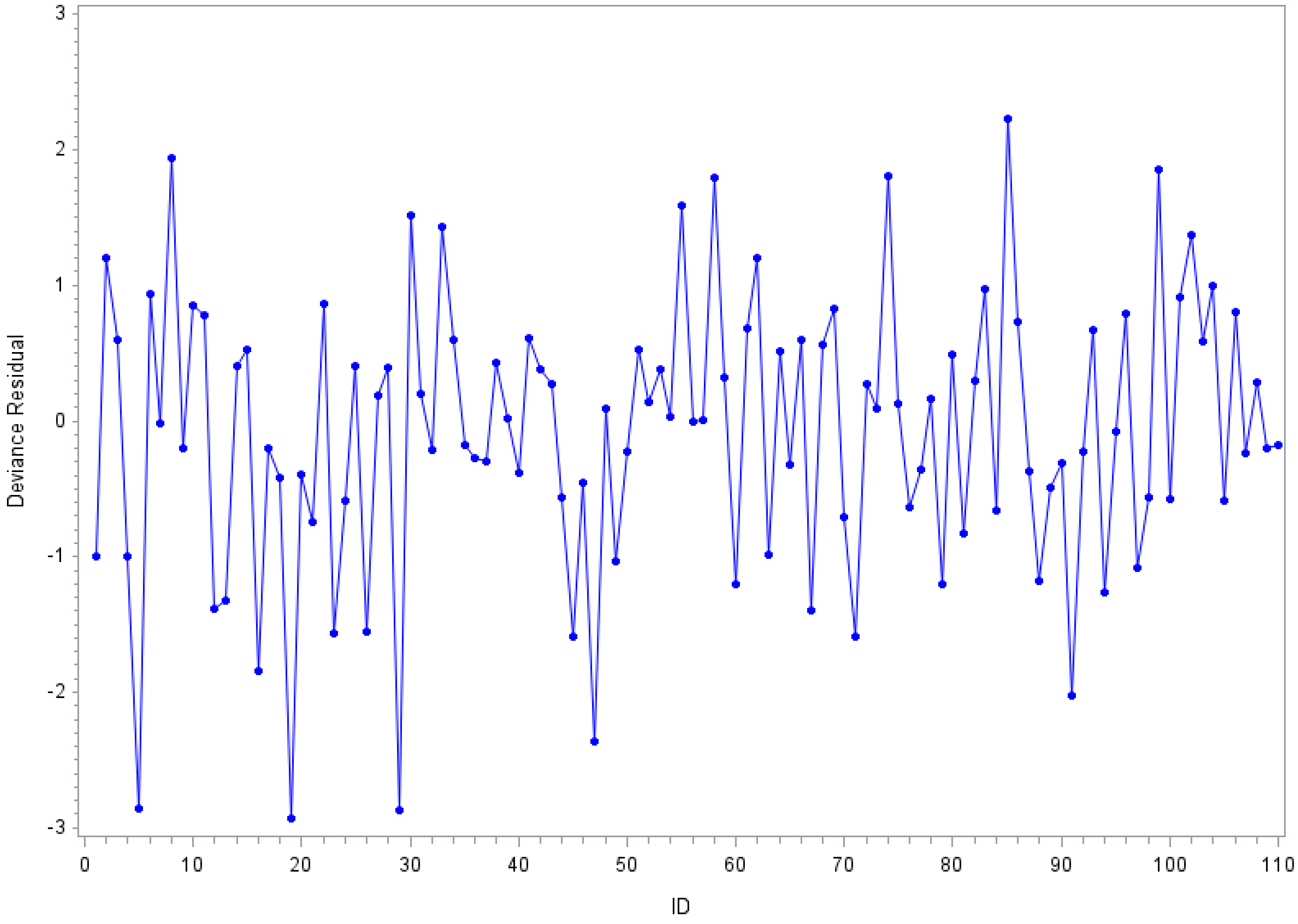
$$Dev_i = \{\text{Sign of } Y_i - \hat{\mu}_i\}[2\, Y_i \log_e(\hat{\mu}_i / Y_i) - 2(Yi - \hat{\mu}_i)]^{1/2}$$

$$Dev_1 = -\,[-(2*9*Log_e(12.3378/9) - 2(9-12.3378)]$$

$$= -\,0.99881$$

We use a similar equation for logistic regression as well.

Deviance Residual Plot



14 End of Lecture 1

The End!